

学校编码: 10384

分类号\_\_\_\_密级\_\_\_\_

学 号: 21620071151909

UDC\_\_\_\_

厦 门 大 学

硕 士 学 位 论 文

**18 种鹭科鸟类 CR1 (Chicken Repeat 1)**

**家族及其与微卫星之间关系的研究**

**Researches of the CR1 (Chicken Repeat 1) family from 18**

**ardeids and its association with microsatellite**

黄志俊

指导教师姓名: 陈小麟 教授

专 业 名 称: 生物化学与分子生物学

论文提交日期:

论文答辩时间:

学位授予日期:

答辩委员会主席:

评 阅 人:

## 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为( )课题(组)的研究成果,获得( )课题(组)经费或实验室的资助,在( )实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

## 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（        ） 1. 经厦门大学保密委员会审查核定的保密学位论文，  
于        年        月        日解密，解密后适用上述授权。

（        ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年        月        日

厦门大学博硕士论文摘要库

# 目 录

摘要.....	1
Abstract.....	3
第一章 前言.....	5
1.1 鸟类基因组重复序列的研究进展.....	5
1.1.1 鸟类基因组中重复序列简介.....	5
1.1.2 长散在重复序列 LINE.....	6
1.1.3 短散在重复序列 SINE.....	7
1.1.4 长末端散在重复序列 LTR-retrotransposons.....	8
1.1.5 转座子 DNA transposons.....	8
1.2 鸟类反转座子 CR1 的研究进展.....	9
1.2.1 CR1 的早期研究.....	9
1.2.2 CR1 的结构.....	11
1.2.3 CR1 的增值模式及家族分类.....	14
1.2.4 CR1 的反转座机制.....	16
1.2.5 CR1 在鸟类及其它物种中的分布.....	17
1.3 CR1 在鸟类分子系统学研究中的应用.....	17
1.4 反转座子和微卫星的关系.....	19
1.5 本研究的目的是和意义.....	22
第二章 材料与方法.....	23
2.1 仪器与试剂.....	23
2.1.1 主要的仪器设备.....	23
2.1.2 试剂.....	23
2.2 样品采集.....	23
2.3 实验方法.....	24
2.3.1.DNA 的提取.....	24
2.3.2 引物设计.....	25

2.3.3 PCR 扩增和产物纯化 .....	27
2.3.4 克隆转化和测序 .....	29
<b>2.4 数据处理与分析 .....</b>	<b>30</b>
2.4.1 鹭科鸟类反转座子 CR1 分析 .....	30
2.4.2 鸟类反转座子 CR1 和微卫星之间的关系分析 .....	32
<b>第三章 结果与分析 .....</b>	<b>33</b>
<b>3.1 鹭科鸟类 CR1 家族的研究结果 .....</b>	<b>33</b>
3.1.1 鹭科鸟类 CR1 序列特征分析 .....	33
3.1.2 鹭科鸟类 CR1 亚家族分类 .....	36
3.1.3 鹭科鸟类 CR1 亚家族在 18 个物种中的分布 .....	38
3.1.4 不同 CR1 亚家族进化年代分析 .....	39
<b>3.2 鸟类 CR1 和微卫星之间的关系研究结果 .....</b>	<b>40</b>
3.2.1 鹭科鸟类部分序列和微卫星之间的结构分析 .....	40
3.2.2 数据库检索 .....	43
3.2.3 鸟类 CR1 和微卫星的关系分析结果 .....	43
<b>第四章 讨论 .....</b>	<b>55</b>
4.1 CR1 序列的富集方法 .....	55
4.2 鹭科鸟类 CR1 序列结构特征 .....	56
4.3 鹭科鸟类 CR1 亚家族分类、进化年代及在不同物种中的分布 .....	60
4.4 鸟类 CR1 和微卫星之间的关系 .....	66
<b>第五章 结论与展望 .....</b>	<b>69</b>
5.1 结论 .....	69
5.2 展望 .....	69
<b>参考文献 .....</b>	<b>71</b>
<b>致谢 .....</b>	<b>78</b>

## Table of contents

<b>Abstract in Chinese</b>	<b>1</b>
<b>Abstract in English</b>	<b>3</b>
<b>Chapter 1 Introduction</b>	<b>5</b>
<b>1.1 Progress of repetitive sequences research of avian genome</b>	<b>5</b>
1.1.1 The repetitive landscape of avian genome	5
1.1.2 Long interspersed nuclear elements (LINEs)	6
1.1.3 Short interspersed nuclear elements (SINEs)	7
1.1.4 Long terminal repeat retrotransposons (LTR-retroposons)	8
1.1.5 DNA transposons	8
<b>1.2 Progress of CR1 investigation</b>	<b>9</b>
1.2.1 The early study about CR1	9
1.2.2 The structure of CR1	11
1.2.3 The propagation model and subfamilies of CR1	14
1.2.4 The retotransposition mechanism of CR1	16
1.2.5 The distribution of CR1 among animals	17
<b>1.3 Application and promise of CR1 methods in avian phylogenic</b>	<b>17</b>
<b>1.4 The microsatellite repeats and retrotransposons</b>	<b>19</b>
<b>1.5 Aims of this study</b>	<b>22</b>
<b>Chapter 2 Materials and Methods</b>	<b>23</b>
<b>2.1 Instruments and reagents</b>	<b>23</b>
2.1.1 Instruments	23
2.1.2 Reagents	23
<b>2.2 Sample collection</b>	<b>23</b>
<b>2.3 Methods</b>	<b>24</b>
2.3.1 DNA extraction	24
2.3.2 Primer design	25
2.3.3 PCR amplify and product purification	27
2.3.4 Clone and Sequencing	29

<b>2.4 Data processing and analysis .....</b>	<b>30</b>
2.4.1 Analysis of the Ardeidae CR1 sequence .....	30
2.4.2 Analysis of association between CR1 and microsatellite .....	32
<b>Chapter 3 Results and analysis .....</b>	<b>33</b>
<b>3.1 Result of Ardeidae CR1 analysis .....</b>	<b>33</b>
3.1.1 Sequence characteristics of Ardeidae CR1 .....	33
3.1.2 Identification of subfamilies of Ardeidae CR1 .....	36
3.1.3 The distribution of Ardeidae CR1 among 18 ardeids .....	38
3.1.4 The evolutionary age of Ardeidae CR1 subfamilies .....	39
<b>3.2 Result of association between CR1 and microsatellite .....</b>	<b>40</b>
3.2.1 Several cases of Ardeidae microsatellite associated with CR1 .....	40
3.2.2 Database search results .....	43
3.2.3 Relationship between CR1 and microsatellite .....	43
<b>Chapter 4 Discussion .....</b>	<b>55</b>
4.1 The enrichment of CR1 sequences .....	55
4.2 Characteristics of CR1 of Ardeidae .....	56
4.3 Phylogenetic of Ardeidae CR1 .....	60
4.4 Intimate association between CR1 and microsatellite .....	66
<b>Chapter 5 Conclusion and Perspective .....</b>	<b>69</b>
5.1 Conclusion .....	69
5.2 Perspective .....	69
<b>References .....</b>	<b>71</b>
<b>Acknowledgements .....</b>	<b>78</b>



## 摘要

本文以 18 种鹭鸟(大白鹭 *Ardea alba*、中白鹭 *A. intermedia*、苍鹭 *A. cinerea*、草鹭 *A. purpurea*、白鹭 *Egretta garzetta*、黄嘴白鹭 *E. eulophotes*、岩鹭 *E. sacra*、池鹭 *Ardeola bacchus*、夜鹭 *Nycticorax nycticorax*、牛背鹭 *Bubulcus ibis*、绿鹭 *Butorides striatus*、海南鴈 *Gorsachius magnificus*、黑冠鴈 *G. melanolophus*、黄苇鴈 *Ixobrychus sinensis*、栗苇鴈 *I. cinnamomeus*、黑苇鴈 *I. flavicollis*、紫背苇鴈 *I. eurhythmus* 和大麻鴈 *Botaurus stellaris*) 为主要研究对象, 通过 PCR 扩增和数据库检索获取 CR1 序列, 探讨两个问题: 第一, 18 种鹭鸟中 CR1 不同亚家族的特征、系统进化及将 CR1 作为分子标记研究鹭科鸟类系统进化关系的可能性; 第二, CR1 和微卫星之间的关系及其对鸟类微卫星开发和应用的影响。主要获得两部分的研究结果:

第一部分通过数据库检索和 PCR 富集两种方法, 18 种鹭科鸟类中共分离到 284 条长度约为 320 bp 的 CR1 序列, 包含 ORF-2 的 3' 末端 267 bp 和 3' UTR 的 48-58 bp。其中 3' UTR 可以分为高变区和保守区, 根据高变区的序列特征, 可以将鹭鸟 CR1 的高变区分为五种类型: Ard-1a, Ard-1b, Ard-2, Ard-3, Ard-4。分析 Ard-1a 和 Ard-1b 高变区的差异, 结合已有的鸟类 CR1 家族数据, 提出在高变区短片段的插入可能是该区域序列多态性的一个产生机制。

利用 40 个已有的 CR1 亚家族构建鸟类 CR1 亚家族间的系统进化关系, 可将鸟类 CR1 分为 5 个世系: Clade1, Clade2, Clade3, Clade4 以及 Clade0, 通过 ORF-2 区段的序列分析, 可以将鹭科鸟类的 CR1 分为 4 个亚家族: Ard-1, Ard-2, Ard-3, Ard-4, 分别对应于鸟类 CR1 的前 4 个世系。利用 Alucode 分析诊断性碱基的原理, 进一步将 4 个亚家族分为 44 个亚亚家族。借用原鸡的核基因中性突变率, 计算出 44 个亚亚家族的进化年代大概在 340 至 5240 万年之间。在此基础上, 参考各个亚亚家族在不同物种中的分布情况, 筛选出 13 个对鹭科鸟类系统进化研究具有潜在价值的亚亚家族。

第二部分通过数据库检索的方法, 在鸟类 9 个目 18 个科的物种中共得到 170 条与 SSR 有关联的 CR1 序列, 其中在 91 条序列里 SSR 和 CR1 紧密相连(间隔 0-18 bp), 另外 79 条序列里的 SSR 则位于 CR1 内部。分析这 79 条序列中 CR1

和 SSR 相毗邻的序列特征，一共可归为三种情况：第一种情况，CR1 含有前微卫星序列，推测其直接参与 SSR 的成熟过程；第二种情况，CR1 含有低复杂度的序列，推测其间接参与了 SSR 的生成；第三种情况，CR1 不含有明显的和 SSR 相关的序列，这部分 SSR 的生成机制尚不清楚。

以 CR1-B 亚家族全长序列为参照序列，SSR 位于 CR1 内部的 79 条序列一共可以定位到 23 个位点，其中有 27 条属于 CR1-X3\_Pass 亚家族的序列在同一位点（3' UTR 末端）生成五碱基重复单元的 SSR，暗示 CR1-X3\_Pass 家族的 3' UTR 末端可能是一个 SSR 进化的热点区域。在余下的 22 个位点中，有 10 个位点发生两例或两例以上的 SSR，另外 12 个位点各有一例 SSR 发生。其余的 CR1 家族是否可以作为 SSR 在基因组中的扩散媒介，还需要更多的数据加以阐明。基于已有的研究中显示的来源于转座元件的 SSR 会对结果分析产生不同程度的影响，建议在鸟类中开发 SSR 位点时要检测其侧翼序列特征。

综上所述，（1）本文首次研究了鹭科鸟类 CR1 亚家族的特征和系统进化，为进一步利用 CR1 作为分子标记研究鹭科鸟类系统进化关系提供了可行性支持，在本研究的基础上，可以设计特异性探针富集带有侧翼信息的 CR1 序列，参考不同亚家族的进化年代，从中选取适用于解析鹭科各属或属内各种之间关系的分子标记；（2）本文首次较系统的研究了鸟类 CR1 和 SSR 之间的关系，发现鸟类中部分 SSR 起源于 CR1 中的前微卫星序列，其中 CR1-X3\_Pass 亚家族对于 SSR 的开发应用是一个值得深入探讨的问题。

**关键词：**鹭科鸟类 CR1 微卫星 系统进化

## Abstract

CR1 fragments were isolated from 18 ardeids (*Ardea alba*, *A. intermedia*, *A. cinerea*, *A. purpurea*, *Egretta garzetta*, *E. eulophotes*, *E. sacra*, *Ardeola bacchus*, *Nycticorax nycticorax*, *Bubulcus ibis*, *Butorides striatus*, *Gorsachius magnificus*, *G. melanolophus*, *Ixobrychus sinensis*, *I. cinnamomeus*, *I. flavicollis*, *I. eurhythmus* and *Botaurus stellaris*) by PCR amplification and database searching. Based on our CR1 dataset, we investigated the structures and the phylogenetic relationships among CR1 subfamilies of 18 ardeids and the association between avian CR1 and SSR elements. The potential of using CR1 as a new phylogenetic marker for ardeids and limitations of applying SSR derived from CR1 were discussed.

284 CR1 fragments were isolated from 18 ardeids through database searching and PCR method. The length of CR1 ranged in size from 315-325 bp which due to the different length of 3' UTR. The hyper-variable domain and conserve domain of 3' UTR are defined based on the sequences alignments. Our data suggested that there are 5 distinct groups of hyper-variable domain within the ardeids, namely, Ard-1a, Ard-1b, Ard-2, Ard-3, Ard-4. In particular, the difference between 3' UTR of Ard-1a and Ard-1b, combining with other previously described avian CR1 subfamilies, prompted speculation that short insertion mutations may be one of the evolutionary mechanisms underlying the diversity of hyper-variable domain.

Phylogenetic analysis of 40 previously described avian CR1 subfamilies indicated that 5 major CR1 clades (clade0 to clade4) may have been present in a common ancestor of most extant bird species. Base on ORF-2 sequences, these ardeids CR1s could divided into 4 distinct subfamilies (Ard-1 to Ard-4) with each corresponding to CR1 clade1 to clade4. Using Alucode method, ardeids CR1s were further characterized into 44 subsubfamilies that share diagnostic nucleotides. The age of subsubfamilies range from 3.4 to 52.4 My were determined from the sequence divergence between every pair of subsubfamily members assuming a substitution rate of  $3.6 \times 10^{-9}$  substitutions/site/year calculated for chicken genome. Finally, 13 potential

phylogenetically informative subsubfamilies were screened out according to the subsubfamily age and their distribution among 18 ardeids.

170 CR1s associated with SSR were identified through searching the database which representing 18 families in 9 avian orders. Among these target sequences, SSR were juxtaposed to CR1 in 91 instances with 0-18 bp gaps between them, while in other 79 instances, SSR were situated precisely in CR1. A detailed examination of the 79 instances revealed that three types of association existed between CR1 and SSR: (i) that CR1 harbored proto-SSR which may produce SSR directly; (ii) that CR1 harbored low complexity sequence which may participate in the production of SSR indirectly; (iii) that CR1 have no obvious characteristic involved in the rise of SSR.

Using CR1-B as a reference sequence, 79 SSRs that located in the CR1 internal sequence could binned to 23 sites. Specifically, 27 SSRs with ATTCT repeat motif or its variants are obviously arose from expansion of a proto-SSR within 3' UTR of CR1-X3\_Pass, indicating that this site may be a hot spot for the rise of SSR. Among the rest of 22 sites, 10 sites have equal or greater than 2 SSRs, and 12 sites have only one SSR. Whether other CR1 subfamilies could act as proto-SSR or SSR generators need more data to elucidate. We suggest that all flanking sequences of SSR coming from library-screening process should be analysed carefully (eg. use Repbase to search repetitive elements in flanking sequences) before using them as polymorphism markers in avian study.

In conclusion, (i) this is the first report for the characterization and phylogenetic analysis among CR1 subfamilies in ardeids, the observation that several recently active subfamilies are present suggests that CR1-based phylogenetic analysis is possible in ardeids. Future work will focus on the development of taxon-specific insertion loci to aid ardeids systematic. (ii) Our data clearly indicate that some types of SSR were evolved from the proto-SSR in CR1, and we hope these efforts will be useful for developing and applying of SSR in avian study.

**Keywords:** Ardeidae; CR1; SSR; Phylogenetic evolution

## 第一章 前言

### 1.1 鸟类基因组重复序列的研究进展

#### 1.1.1 鸟类基因组中重复序列简介

基因组中的重复序列也被称为“垃圾序列”(junk DNA)或者“自私因子”(selfish DNA),包括串联重复序列(short tandem repeats)以及各种类型的散在重复序列(interspersed repeats)<sup>[1]</sup>。散在重复序列又称为转座因子(mobile elements),根据转座机制的不同可以分为两类:第一类以 mRNA 为中介,称为反转座子;第二类是直接以 DNA 的形式转座,称为转座子。反转座子可以根据是否含有长末端重复(LTR)分为两类:长末端反转座子(LTR-retrotransposons)以及非长末端反转座子(non-LTR-retrotransposons)。其中第二类根据其序列长度可以再细分为两类:短散在重复序列(SINE, Short Interspersed Nuclear Element)和长散在重复序列(LINE, Long Interspersed Nuclear Element)。

不同生物类群的基因组中重复序列所占的比例存在较大的差异。2004 年测得的原鸡(*Gallus gallus*)基因组大小为 1200 Mbp<sup>[2]</sup>,只占人类基因组大小的 35%,老鼠基因组大小的 45%。如此显著的差异很难从功能基因的数目方面得到解释,因为用不同的预测功能基因的方法得到的原鸡中功能基因的数目大概为 20000-23000 个,和哺乳动物并无显著差异。比较基因组学研究和重组动力学研究发现,基因组中重复序列的含量是造成这一差异的一个关键因素:哺乳类基因组中含有 40%-50%的重复序列,而鸟类则只有 10%<sup>[3-5]</sup>。

鸟类的染色体中多数为小染色体(microchromosomes, 0.5-2.5  $\mu\text{m}$ ),只有少数几对为大染色体(macrochromosomes, 3-6  $\mu\text{m}$ ),如原鸡有 33 对小染色体<sup>[6]</sup>,而大染色体只有 6 对。小染色体普遍存在成对交叉的现象,有利于有丝分裂和减数分裂时小染色体之间的配对<sup>[7]</sup>。Burt 等<sup>[8]</sup>认为鸟类基因组中较少的重复序列也有利于小染色体同源区域的配对。Hughes 等<sup>[9]</sup>比较了原鸡和人类不同大小的染色体中重复序列的比例,发现在染色体长度相似的情况下,人类的染色体中重复序列比例远远高于原鸡的,同时,原鸡染色体中所含的重复序列随着染色体长度的增加而增加。

### 1.1.2 长散在重复序列 LINE——CR1

原鸡的基因组研究结果表明,有 9%的序列属于散在重复序列,其中数量最多的是一类称为 CR1 (Chicken Repeat 1) 的非长末端反转座子,拷贝数达到 20 万个,占基因组中散在重复序列的 80%以上(表 1-1)。完整的 CR1 全长有 5-6 kb,但大多数的 CR1 都在 5' 端发生不同程度的缺失,只剩 3' 端几百 bp 的序列。这暗示了 CR1 的 3' 末端与其余的片段相比可能存在某种调节功能<sup>[10]</sup>。但也有学者指出冗余的 3' 端可能仅仅是 CR1 转座过程过早结束 (premature termination) 的产物,由于 5' 端编码与反转座相关的重要的酶,这些残缺的 CR1 在插入基因组位点时便不再有转座功能 (dead-on-arrival)<sup>[11]</sup>。原鸡的基因组中只找到一个完整的具有编码功能的 CR1,另外一些较长的 CR1 序列中普遍存在碱基突变或者插入缺失突变,使得编码区丧失功能。由于大多数 CR1 都是没有功能的,推测原鸡的基因组中可能已经不存在有活性的 CR1<sup>[12]</sup>。另外,原鸡的基因组中存在多个不同的 CR1 家族,系统进化分析表明,大部分 CR1 在 4500 万年前曾到达一个活性的峰值,之后便逐渐沉默<sup>[12]</sup>,一些更为古老的 CR1 家族甚至在鸟类和爬行类分化之前就已经存在<sup>[13, 14]</sup>。

表 1-1 原鸡基因组中散在重复序列的类型、比例及其分布

(International Chicken Genome Sequencing Consortium, 2004)

Tab 1-1 Composition of interspersed repeats in the chicken genome

Repeat type	Copy number	Density				
		Overall	Macro	Micro	Z	Unassigned
		(%)	(%)	(%)	(%)	(%)
CR1	205,000	6.4	7.4	3.2	10.4	8.0
MIRs/LINE2	10,000	0.1	0.1	0.1	0.1	0.1
LTR elements	12,000	1.3	1.3	0.5	1.8	3.3
DNA transposons	13,000	0.8	1.0	0.3	1.5	0.8
Simple repeats	12,000	0.7	0.6	0.4	0.7	1.5
Satellites	2,000	0.1	<0.1	<0.1	<0.1	0.9
Total	254,000	9.4	10.2	4.5	14.5	14.5

原鸡基因组中具有编码功能的序列只有 4%，除去 10%左右的重复序列，还有超过 85%的序列（200 Mb）目前仍然没有得到充分的研究。基于反转座子的古老起源，推测在这未破解的序列中，有相当大的一部分也属于散在重复序列，由于它们在基因组当中经历了漫长的演化，已经很难通过序列比对来识别它们，因此已有的关于 CR1 拷贝数的估计应该是保守的，实际数目或许要远大于 20 万个<sup>[15]</sup>。通常认为重复序列不受选择压力的作用，因此它们可以在基因组中通过积累突变而产生高度的复杂区域，这些区域有可能是功能基因的起源之一<sup>[5]</sup>。另外，某些家族的成员之间的序列差异度很大，在暗示其古老起源的同时，也提示原鸡基因组当中较少的散在重复序列并不是由于基因组高频率的缺失造成的，更可能是重复序列在原鸡进化过程中逐渐沉默的结果<sup>[2]</sup>。

### 1.1.3 短散在重复序列 SINE

SINE 是与 LINE 密切相关的一类散在重复序列。它是基因组中的一种中度重复序列，长度在 70–500 bp 之间<sup>[16]</sup>，在真核生物基因组中的拷贝数大约在  $10^3$ – $10^5$  之间<sup>[17]</sup>，它已经在人类<sup>[18]</sup>、哺乳类<sup>[19]</sup>、两栖类<sup>[20]</sup>、鱼类<sup>[21]</sup>、真菌<sup>[22]</sup>和植物<sup>[23]</sup>等许多真核生物中发现。SINE 的结构一般由三部分组成：5' 端的 RNA 相关区，这部分衍生于相应的 7SLRNA、tRNA 或者 5SrRNA；中部的 RNA 非相关区，这部分变异很大，而且没有特征性的模式；3' 端的 AT 富含区，有时也称为 LINE 衍生区<sup>[24]</sup>，该区最显著的特征是含有丰富的 AT，与 LINE 的 3' 端尾部有很高的相似性。因为 LINE 编码的反转座酶能无特异性的识别 3' 端的 AT 富含区来介导 LINE 的反转座，所以在所有关于 SINE 的研究中，都把 SINE 的转座看作是相应的 LINE 相关联的，SINE 也因此被称为 LINE 的寄生因子<sup>[25, 26]</sup>。由于 SINE 和 LINE 的密切关联，理论上原鸡基因组中存在的大量的 CR1 暗示着大量相关的 SINE 的存在，但是在已测得的原鸡的基因组中并没有发现与 CR1 相关联的 SINE，这被认为是与 CR1 编码的反转座酶的特异性有关<sup>[2]</sup>。如上所述，哺乳动物的 L1 (LINE-1) 编码的反转座酶可以识别大量的富含 AT 尾巴的序列，而 CR1 编码的反转座酶高度特异的识别 3' 端八碱基重复，因此该酶无法被其余的以 Poly-A 结尾的散在重复序列所利用，是 CR1 自身转座专一性的酶。同样，由于 L1 和 CR1 各自编码的反转座酶的特异性不同，也导致相应基因组中假基因的数

量差别：原鸡的基因组中只检测到 51 个由转座因子介导形成的假基因，而哺乳动物的这类假基因则多达 15000 个<sup>[27]</sup>。

在原鸡的基因组中检测到的 SINE (MIR 和 MIR3, 分别与 L2 和 L3 相关联) 都是古老且丧失活性的, 研究表明这些 SINE 在鸟类和哺乳类分化之前就已经存在, 但在不同的支系中它们的命运却显著不同: 在哺乳动物中, 这些古老的 SINE 逐渐被新生的 SINE 所掩盖甚至替代, 而在鸟类中大多数都被完整的保留<sup>[9, 12]</sup>。

#### 1.1.4 长末端散在重复序列 LTR-retrotransposons

这类重复序列元件的典型特征是在中央编码区域的两侧带有很长的侧翼重复序列, 这些侧翼序列通常较长 (几百 bp), 并且含有反转座过程必须的一些调节因子 (如启动子)<sup>[28, 29]</sup>。LTR-retrotransposons 的结构和反转录病毒的结构很相似, 通常认为它起源于反转录病毒。两者都含有 *gag* 和 *pol* 基因, 前者编码病毒衣壳蛋白, 后者编码反转座过程必须的酶, 如反转录酶, 核酸内切酶<sup>[1]</sup>。与其他的脊椎动物相类似, 原鸡基因组中 LTR-retrotransposons 的数量要远远的小于 non-LTR-retrotransposons。通过比较原鸡基因组中的 class II GGERVK10 和 class III GGERVL 的序列相似性后发现它们各自家族成员之间的差异度只有 3%, 推测也许目前仍然处于活跃时期。在哺乳动物中, 有一类和 GGERVL 相对应的重复序列, 称为 ERV-L<sup>[30]</sup>, 研究表明这类元件有着古老的起源, 因此推测原鸡当中的 GGERVL 和哺乳动物中的 ERV-L 有共同的祖先, 并且在哺乳类和鸟类分化前就已定居在基因组中。但是这和观察到的 GGERVL 家族成员间的序列差异度是矛盾的, 也许 GGERVL 和 ERV-L 是在鸟类和哺乳类各自进化过程中反转录病毒独立渗透的产物<sup>[2]</sup>。

#### 1.1.5 转座子 DNA transposons

与上述的几类反转座子显著不同的是, 反转座子采用 “copy and paste” 的转座策略, 而转座子的转座策略为 “cut and paste”, 即一个转座子从基因组中的某一位点将自身完整的切下, 然后再插入到另一个新的位点<sup>[31]</sup> (图 1-1)。

原鸡基因组中转座子 DNA 数目较少, 相关的研究也不多, 目前已发现的转座子都属于两个家族: Charlie12\_GG (activator-like) 和 GGMAR (mariner-like)。推



Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库